Evaluating Mobile and Ubiquitous Applications in the Field by Automated Capture and Analysis of Reality Traces

Jensen, Kasper Løvborg

Published in:
Proceedings of the Eighth Danish Human-Computer Interaction Research Symposium, 20th November 2008

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Evaluating Mobile and Ubiquitous Applications in the Field by Automated Capture and Analysis of Reality Traces

Kasper Løvborg Jensen
Multimedia Information and Signal Processing
Department of Electronic Systems
Aalborg University
klj@es.aau.dk

ABSTRACT
New methods and tools are needed to face the challenges of evaluating the user experience of mobile and ubiquitous applications. The paper discusses reality traces as the results from capturing information about the interaction with such applications and the context in which it occurs. Automated capture can be done by utilizing the sensing and processing power of personal mobile devices and be used to conduct remote and autonomous field evaluations in realistic settings. The paper presents RECON and GREATDANE, which are two concrete tools developed for respectively capturing and analyzing such reality traces.

Keywords
Mobile, ubiquitous, user experience, interaction, context, reality traces, data capture, data analysis, method, tools.

1. INTRODUCTION

Innovative mobile applications are emerging powered by advances in technology and computing paradigms such as ubiquitous, pervasive and context-aware computing. Interacting with such applications often entails a more complex user experience that requires special attention when evaluating it.

While UX is generally used as shorthand for user experience, this paper will use the term µX (MUX) to refer specifically to the Mobile and Ubiquitous user eXperience. µX can be defined as *"the user experience arising from systems, services and applications with which the interaction is essentially mobile and ubiquitous"* [14].

Mobility should be attributed both to the device and the user [9], and together with the situatedness it gives rise to complex and unpredictable contexts of use which will directly or indirectly influence the user's experience.

Evaluating µX

A central discussion concerns whether to evaluate applications in the laboratory or in the field. Intuitively, µX should be studied in-situ under realistic conditions, yet this is not the predominant approach [16]. Arguments against field studies are that that data collection is difficult, costly and that such experiments lack control of the contextual parameters. Some claim that it is not worth the hassle [17, 15] and others that it is [19].

Some key problems with existing methods are:

- The obtrusiveness to the user experience as they rely on the user to actively report data or observers to be physically present.
- The time, manpower and resources needed to design, set up and conduct the experiments are high.
- They do not scale well with the number of users, duration of study, and geographic area in which they can be conducted.
- The lack of capability to study long term usage and/or interaction in context properly.

The purpose of this paper is to discuss remote and autonomous field evaluations as a new approach for evaluating µX applications that address the abovementioned problems.

Paper Outline

Section 2 will introduce the concept of reality traces and discuss the importance of context. Section 3 will discuss how field studies for gathering such data can be facilitated through automated capture and analysis of reality traces and Section 4 and 5 will present two concrete tools for doing this. Section 6 will discuss pros and cons of doing this type of studies.

2. REALITY TRACES

Reality traces are datasets describing the users' interaction with an application and the context in which it occurred. Essentially it is detailed log files augmented with contextual information about the particular situations.

Interaction Data

Interaction can be considered at various levels of abstraction. From low level UI events like button presses to higher level actions, activities and sessions. To get a detailed picture of the user experience it might be necessary to consider all of these levels. Reality traces should contain all information of relevance in the later analysis.

Context

Context is very important to µX and thus also a critical factor to consider when evaluating it. A frequently cited definition for context within context-aware computing is given by Dey:

> *"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." [4]*

The notion of relevancy is not very clear and also the above definition is for context-aware systems, i.e. systems that *"…uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task"* [4]. When discussing evaluation of µX the following definition is proposed instead, which is essentially based on the above but emphasizes the user experience and the situatedness:

> *"Context is the sum of relevant factors that characterize the situation of a user and a system, where relevancy implies that these factors have significant impact on the user's experience when interacting with that system in that situation."*

This definition is arguably still prone to the critique of Dourish, who points out that treating context like a representational problem is not realistic since it is not a static concept that can be neatly captured, modeled and represented [5]. Due to the dynamic nature of context a factor may be relevant in one instant of time and irrelevant the next, just as the significance to the user experience may change depending on the situation.

When talking about capturing context for creating reality traces, it thus refers to capturing information about these factors; while acknowledging that it will never be complete and it will be up to the individual evaluators to specify which are of relevance for their specific studies.

3. REMOTE AND AUTONOMOUS FIELD EVALUATION

Figure 1 shows a conceptual diagram of how such field evaluations can be conducted in an unobtrusive way. Capturing software is installed on a mobile device together with the given application. The users in the experiment will interact with the application in their natural environments for a period while the evaluator is spatially and temporally remote. Reality traces are automatically captured and reported to a central server where they can be analyzed and reviewed by the evaluator during the study. Some experiment control is possible through remote configuration of the capturing software.

4. RECON

RECON (made up of Remote and CONtext) is a tool for capturing application specific interactions, general usage of the device and a wide range of contextual factors of the device, e.g. available networks, GSM towers and signal strength, Bluetooth devices within proximity, battery status, etc. [18].

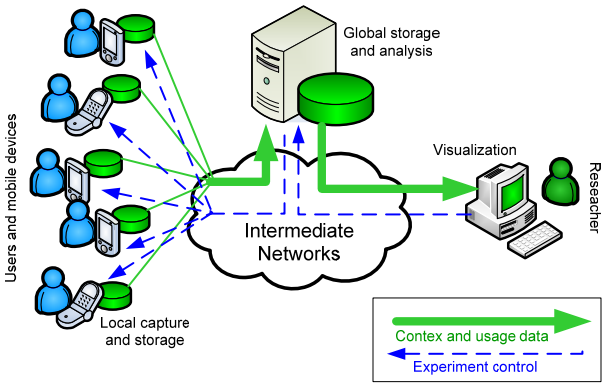

Figure 1: Remote and Automated Evaluation

Figure 2 shows the client-server architecture of RECON. Interaction data is captured through code instrumentation of the application being evaluated. Relevant user and system actions are sent from the application to the RECON client, where they are logged together with the context data in a local database. The context is captured through the sensors and information processes available on the device. The reality traces are kept in local storage on the device until preset conditions are reached and it can be uploaded to a central RECON server. Configurations are updated through the server e.g. the logging policy.

Existing work

Log files is an old and widely used technique, but only few other systems exists for capturing context from the mobile personal devices. The following have all been used in field studies: ContextPhone [21] used in [6] and [20], MyExperience [8] used in [2] and SocioXensor [11] used in [12]. Some have made external sensing devices, which require the user to carry and extra device [1]. RECON is mostly comparable to MyExperience and SocioXensor as they have similar functionalities and run on the same platform (Windows Mobile 5.0 or newer).

The main difference is in how they are set up and deployed and the fact that RECON is especially tuned to capturing detailed interaction data from an application together with context. If needed, RECON can also be used for general usage and/or context capture and it supports prompting the user with small questionnaires.

5. GREATDANE

GREATDANE (Generic REAlity Traces Data ANalysis Engine) is, as implied by the name, a generic tool for analyzing the data contained in a captured set of reality traces. In essence, the goal is to transform these traces into meaningful concepts and metrics from which the user experience can be evaluated. As datasets quickly become large when sampling the context of use, it is very desirable

to automate the processing of these. Thus automation is important to ensure scalability with regard to duration and number of users.

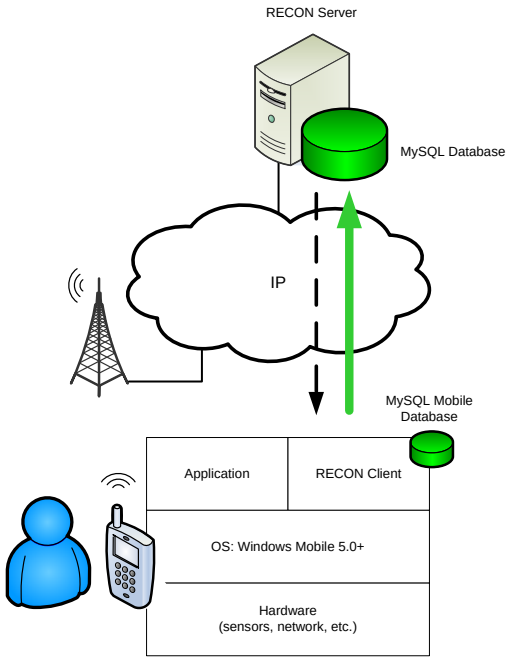

Figure 2: RECON

The analysis builds on a sequential event-based model wherein interaction is treated as a sequence of events categorized as either user actions, system actions or contextual events. The analysis engine uses a predefined model of the application to abstract low-level events and actions into higher level concepts, e.g. sessions, activities, actions. Information about frequency of use, duration, success and error rates and other metrics are calculated at each level.

The approach can in many ways be compared to LSA (Lag Sequential Analysis) as used in [3] and is also inspired by Exploratory Sequential Data Analysis [7]. Hilbert and Redmiles surveyed a range of existing methods for extracting usability information from user interface events [10], but none of the tools and methods there applies directly to µX.

First experiences with GREATDANE were from the DiasNet Mobile field evaluation. Results from the longitudinal field study of DiasNet Mobile can be found in [13].

GREATDANE is designed to be a generic tool independent of the way in which the reality traces are captured. Thus ContextPhone, MyExprience, SocioXensor or any other capturing tool could be utilized to obtain the dataset. It is very important to address the representational problem of reality traces and how it can be synthesized with many users, many contexts, heterogeneous platforms, sensors, etc. If a rich, flexible and uniform representation format of such data was developed and agreed upon, it would enable sharing of datasets and pave the way for generic analysis and data visualization tools.

6. DISCUSSION

The following lists some of the main pros and cons in doing this kind of experiments:

Pros

- Unobtrusive, ideally the user will forget that he/she is in a test situation
- Real tasks, not some scripted scenarios
- Real context, the situations in which the usage will occur is realistic
- Scalability both with regard to area of study, number of users and duration of study
- No need for external sensors

Cons

- Lack of direct control during the experiment
- Noise and uncertainties in the data
- Worst case the users will not use it at all
- Need a working prototype – robust enough for deployment without too much maintenance
- Privacy, security and ethical issues

The objective nature of reality traces can be complimented by qualitative data collected with other methods such as interviews, surveys, etc. One method that fits well with the autonomous field evaluation approach is experience sampling. Which RECON already supports.

Conclusions

New methods are needed to evaluate the mobile and ubiquitous user experience, especially for investigating long term usage and interaction in context. The paper introduced the concept of reality traces and discussed how such data can be captured unobtrusively in remote and autonomous field evaluations using tools such as RECON and GREATDANE for automated capture and analysis.

ACKNOWLEDGMENTS

The author wishes to thank Tais Holland Mogensen and Christian Ølholm for their work on RECON as part of their Master's Thesis at Aalborg University [18].

REFERENCES

1. Choudhure, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca A., LeGrand L., Rahimi A., Rea, A., Borriello, G., Hemingway, B., Klasnja, P., Koscher, K., Landay, J. A., Lester, J., Wyatt, D., & Haehnel, D. The mobile sensing platform: an embedded activity

recognition system. IEEE Pervasive Computing, 7(2):32-40, 2008.

2. Consolvo, S., McDonald D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., & Landay, J. A. Activity sensing in the wild: a field trial of ubifit garden. In Proceedings of CHI 2008.
3. Consolvo, S., Arnstein, L., & Franza, B. R. User study techniques in the design and evaluation of a ubicomp environment. In proceedings of the 4th international conference on Ubiquitous Computing, 2002.
4. Dey, A. K. Understanding and using context. Personal Ubiquitous Computing, 5(1):4-7, 2001.
5. Dourish P. What we talk about when we talk about context. Personal Ubiquitous Computing, 8(1):19-30, 2004.
6. Eagle, N. and Pentland. A. S. Reality mining: sensing complex social systems. Personal Ubiquitous Computing, 10(4):255-268, 2006.
7. Fisher, C. & Sanderson, P. Exploratory sequential data analysis: exploring continuous observational data. Interactions, 3(2):25-34, 1996.
8. Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., & Landay, J. A. Myexperience: A system for in situ tracing and capturing of user feedback on mobile phones. In Proceedings of MobiSys 2007.
9. Gorlenko, L. & Merrick, R. No wires attached: Usability challenges in the connected mobile world. IBM Systems Journal, 42(4):639-651, 2003.
10. Hilbert, D. M. & Redmiles, D. F. Extracting usability information from user interface events. ACM Computing Surveys, 32(4):384-421, 2000.
11. ter Hofte, G. H What's that Hot Thing in my Pocket? SocioXensor, a smartphone data collector. In Proceedings of e-Social Science 2007.
12. ter Hofte, G. H. Xensible interruptions from your mobile phone. In Proceedings of MobileHCI 2007.
13. Jensen, K. L. & Larsen, L. B. Evaluating the usefulness of mobile services based on captured usage data from longitudinal field trials. In Proceedings of Mobility 2007.
14. Jensen, K. L. & Larsen, L. B. The challenges of evaluating the mobile and ubiquitous user experience. In Proceedings of IMUx 2008.
15. Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. Usability testing of mobile applications: A comparison between laboratory and field testing. Journal of Usability Studies, 1(1):4-17, 2005.
16. Kjeldskov, J. & Graham, C. A review of mobilehci research methods. In Proceedings of Mobile HCI 2003.
17. Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In Proceedings of Mobile HCI 2004.
18. Mogensen, T. H. & Ølholm, C. RECON: A framework for remote automated field evaluation of mobile systems. Master's thesis, Aalborg University, 2007.
19. Nielsen, C. M., Overgaard, M., Pedersen, M. B., Stage, J., and Stenild, S. Its worth the hassle! the added value of evaluating the usability of mobile systems in the field. In Proceedings of NordiCHI, 2006.
20. Oulasvirta, A., Petit, R., Raento, M., & Tiita, S. Interpreting and acting on mobile awareness cues. Human-computer Interaction, 21:97-135, 2007.
21. Raento, M., Oulasvirta, A., Petit, R., & Toivonen, H. ContextPhone - a prototyping platform for context-aware mobile applications. IEEE Pervasive Computing, 4(2):51-59, 2005.